ON THE SUFFICIENCY AND LIKELIHOOD PRINCIPLES

D. A. S. Fraser
University of Toronto

The likelihood, sufficiency, and conditionality principles proposed by Allan Birnbaum are discussed in the context of two examples that include in their specifications a position relationship between the variable and the parameter. The discussion leads to the suggestion that the likelihood principle is too strong and that the sufficiency and conditionality principles should be used only when the necessary analysis is consistent with the position relationship between variable and parameter.

I. INTRODUCTION

In a recent paper [1] Allan Birnbaum discusses the roles of sufficient statistics, ancillary statistics, and likelihood functions in informative statistical inference. He proposes three principles related to these concepts and demonstrates an interesting relationship among them. The concepts and an interpretation of the principles may be described briefly as follows:

Let E designate the statistical model for an observable variable x, and let θ be the parameter in this model. A statistic $t(x)$ is a sufficient statistic if the conditional distribution given t does not depend on θ. This independence of the parameter is the justification for a reduction in which inferences are made from the observed value of the statistic $t(x)$ rather than from the original variable x. The principle being used is the following:

Sufficiency Principle (S): In the presence of a sufficient statistic $t(x)$ with statistical model E' the inferences concerning θ from E and x should be the same as from E' and $t(x)$.

The use of ancillary statistics has only recently been receiving attention outside the writings of R. A. Fisher. A statistic $a(x)$ is ancillary if it has a fixed distribution regardless of the parameter value and in some contexts if it depends on x only through an exhaustive statistic.‡ Fisher [3] and Cox [2] have presented examples in which inferences concerning the parameter are made in the framework of the conditional model given the value of the ancillary statistic, the justification being that the ancillary statistic describes some aspect of the structure of the sample and that this aspect of structure should in a sense be accepted and inferences based on a comparison with other samples having this same structure.

Conditionality principle (C): In the presence of an ancillary statistic $a(x)$ having a conditional statistical model E_a given that $a(x) = a$, the inferences concerning θ from E and x should be the same as from $E_{a(x)}$ and x. This is a stronger principle; a simple argument shows that the sufficiency principle can be deduced from it.

The likelihood function has a prominent place in the writings of R. A. Fisher. Let $f(x|\theta)$ be the probability function for x given θ. The likelihood function

‡ A sufficient statistic that provides the maximum reduction of an observable variable is often called a *minimal sufficient statistic*, but is called an *exhaustive statistic* by R. A. Fisher.

$L(\theta|x)$ of θ from a possible outcome x is the probability function $cf(x|\theta)$ treated as a function of θ and left indeterminate to an arbitrary constant multiplier. When the strong forms of inference provided by the Bayes and fiducial methods are not available, Fisher suggests that inferences be based on the likelihood function. As the likelihood function in its dependence on x is a minimal sufficient statistic, this procedure if accompanied by a reference to the statistical model E would be in accordance with the sufficiency principle. The likelihood principle, however, is apparently stronger and requires that the likelihood function be used without reference to the statistical model.

Likelihood principle (L): The inferences concerning θ from E and x should be the same as from $L(\theta|x)$.

Birnbaum shows that the likelihood principle is logically equivalent to the conditionality principle. The conditionality principle does, however, have greater acceptance among statisticians concerned with informative inference. Part of this acceptance could certainly be based on experience with inference in the sciences, external to the logic of statistical theory. Perhaps then the usual statistical model may not be incorporating all the relevant features for informative inference. In the usual model the only connection between variable and parameter is that provided by the probability function $f(x|\theta)$. Thus it would seem reasonable to qualify the likelihood principle by a notation that any force it derives from the conditionality principle is based on a proof that assumes there is no relationship between x and θ other than is provided by the probability function $f(x|\theta)$.

There would seem to be some good grounds for saying that a *general* statistical model was superficial if it admitted no variable-parameter relationship other than that provided by the probability function. And yet such a general model pervades statistical theory. An exception is the use of transformations and invariance which implicitly supply a variable-parameter relationship. A related exception is the analysis of fiducial inference in references [4] and [5] which argues from an error variable relating x to θ; in this, the error variable provides a position relationship between the variable and the parameter.

In this paper several examples are used to study the effect of adding a position relationship between an observable variable and a parameter. The examples suggest that the sufficiency and likelihood principles are too strong for use without qualification and appraisal against the position relationship.

2. A DISCRETE EXAMPLE

Let θ be a parameter that can take any integral value, and suppose there is an instrument for measuring θ that produces an unbiased measurement x with the following distribution:

$$\begin{aligned} P(x \mid \theta) &= \tfrac{1}{5} \qquad \text{if } x = \theta - 2, \theta - 1, \cdots, \theta + 2 \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

For the observation $x=10$, the likelihood function assigns equal values to $\theta=8, 9, \cdots, 12$ and zero value elsewhere. The discussion in Birnbaum's paper suggests various interpretations that can be given to this likelihood function depending on the kind of inference wanted.

This model is translation-invariant in a way that would typically be meaningful for a measuring instrument. The invariant-fiducial analysis in references [4] and [5] is applicable, and produces a probability distribution for θ given $x=10$:

$$\begin{aligned} P(\theta \mid x = 10) &= \tfrac{1}{5} \qquad \text{if } \theta = 8, 9, \cdots, 12 \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

The analysis in references [4] and [5] also supplies a frequency interpretation for this distribution.

The simplicity of this model admits an easy and direct frequency interpretation. Whenever the instrument is used there is an equal probability that the measurement will be -2 units, -1 unit, $\cdots$, $+2$ units from the value of θ; thus the relative frequencies from past and possible measurements will be equal for the five positions $-2, -1, \cdots, +2$ of the observation with respect to the parameter, and also equal for the equivalent positions $+2, +1, \cdots, -2$ of the parameter with respect to the observation. But the model is translation-invariant and each x value enters symmetrically with respect to any other x value. The frequency distribution of parameter position when associated with the particular value $x=10$ then produces the distribution presented in the preceding paragraph. *This distribution is* thus *the frequency distribution of parameter position as translated, on the grounds of the symmetry in the model, to the particular outcome* $x=10$.

Consider now a second instrument for measuring θ that produces a variable y having greater precision, but with a one-unit positive bias for $\theta<10$ and a one-unit negative bias for $\theta>10$; suppose its distribution for $\theta<10$ is

$$\begin{aligned} P(y \mid \theta) &= \tfrac{1}{3} \qquad \text{if } y = \theta, \theta + 1, \theta + 2 \\ &= 0 \qquad \text{otherwise,} \end{aligned}$$

for $\theta=10$ is

$$\begin{aligned} P(y \mid \theta) &= \tfrac{1}{3} \qquad \text{if } y = 9, 10, 11 \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

and for $\theta>10$ is

$$\begin{aligned} P(y \mid \theta) &= \tfrac{1}{3} \qquad \text{if } y = \theta - 2, \theta - 1, \theta \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

For the observation $y=10$, the likelihood function assigns equal value to $\theta=8, 9, \cdots, 12$, and zero value elsewhere. This is the same likelihood as was obtained from $x=10$. The inferences then, according to the likelihood principle must be the same.

Let $\phi=\theta+1$ for $\theta<10$, $\phi=10$ for $\theta=10$, and $\phi=\theta-1$ for $\theta>10$. The distribution of y can be expressed simply in terms of ϕ,

$$\begin{aligned} P(y \mid \phi) &= \tfrac{1}{3} \qquad \text{if } y = \phi - 1, \phi, \phi + 1 \\ &= 0 \qquad \text{otherwise,} \end{aligned}$$

and the parameter ϕ *indexes* the distributions whereas two values of θ could correspond to the same distribution.

In the second form the distribution is translation-invariant and the invariant fiducial analysis can be applied. The resulting fiducial distribution from the measurement $y=10$ is

$$\begin{aligned} P(\phi \mid y = 10) &= \tfrac{1}{3} \qquad \text{if } \phi = 9, 10, 11 \\ &= 0 \qquad \text{otherwise,} \end{aligned}$$

and in terms of the parameter θ it is

$$\begin{aligned} P(\theta \mid y = 10) &= \tfrac{1}{3} \qquad \text{if } \theta = 8, 12 \\ P(\theta = 9, 10, \text{ or } 11 \mid y = 10) &= \tfrac{1}{3}. \end{aligned}$$

This distribution recognizes implicitly that the three parameter values 9, 10, 11 are equivalent as far as the observable variable y is concerned.

This model also has the simple features that permit a direct interpretation of the posterior distribution. The relative frequencies from past and possible measurements will be equal for the three positions $+1$, 0, -1 of the parameter with respect to the observation. Locating this position distribution, on the grounds of translation symmetry, at the value $y=10$ produces the probability distribution in the preceding paragraph.

In this example the likelihood function is the same for the two instances. And yet recognition of an implicit position relationship in the measurement process indicates that different inferences are appropriate to the two cases. This suggests that the likelihood principle as a prescriptive guide to inference should be qualified when there is a meaningful position relationship between the variable and the parameter.

3. A CONTINUOUS EXAMPLE

For a second example let θ be a real valued parameter and suppose that an observable variable x is normally distributed with mean θ^3 and variance 1. Also suppose that the variable x has position with respect to θ^3 which has the same translation invariance as the distribution form. The likelihood function for an observation x is

$$\exp\left\{-\tfrac{1}{2}(x - \theta^3)^2\right\}$$

and for the observation $x=0$ is

$$\exp\left\{-\tfrac{1}{2}\theta^6\right\}.$$

Consider now a second variable y with the following probability function

$$f(y \mid \theta) = k \exp\left\{-\tfrac{1}{2}(y - \theta)^6\right\}$$

and suppose that y has position with respect to θ with the same translation-invariance as the distribution form. The likelihood function from the observation $y=0$ is

$$\exp\left\{-\tfrac{1}{2}\theta^6\right\},$$

which is the same as for $x=0$. The likelihood principle would require the inferences concerning θ from $y=0$ to be the same as from $x=0$.

The statistical model for the first variable is translation-invariant in terms of x and θ^3. The invariant fiducial argument is thus applicable and it yields the following posterior distribution for θ given x

$$k \exp\left\{-\tfrac{1}{2}(\theta - x)^3\right\} d\theta^3,$$

which for $x=0$ becomes

$$k \exp\left\{-\tfrac{1}{2}\theta^6\right\} \theta^2 d\theta.$$

On the basis of the translation-invariance of position and probability function, this describes in a well defined sense [5] the frequency distribution of possible parameter values.

The second statistical model is also translation-invariant but in terms of the variables y and θ. The posterior distribution for θ given y is

$$k \exp\left\{-\tfrac{1}{2}(y - \theta)^6\right\} d\theta,$$

which for $y=0$ becomes

$$k \exp\left\{-\tfrac{1}{2}\theta^6\right\} d\theta.$$

This distribution also has a frequency interpretation on the basis of the translation-invariance of position and probability function.

These posterior distributions are closely related to the likelihood function. For the second variable, the posterior density function is just the likelihood function; for the first variable, the posterior density function is the likelihood function as modulated by the factor θ^2 which depends only on the parameter. Thus the posterior density function as it depends on the observable variable has the structure of a minimal sufficient statistic and its use is in conformity with the sufficiency principle.

However, the recognition of the position relationship and its utilization in the translation-invariance leads to different inferences for the two instances $x=0$ and $y=0$. This is contrary to the likelihood principle and suggests that the likelihood principle is too strong when there is a meaningful position relationship between the variable and the parameter.

4. ON PRINCIPLES OF INFERENCE

The examples in Sections 2 and 3 indicate that the likelihood principle may be too strong in the presence of relationships between the variable and the parameter beyond the minimum embodied in the probability function. Birnbaum's argument mentioned in Section 1 demonstrates that the likelihood principle is equivalent to the conditionality principle. The examples thus show that, in the presence of additional relationships between the variable and the parameter, the conditionality principle is too strong, or the Birnbaum argument is no longer valid, or both.

Consider further the discrete example of Section 2. Suppose that with probability $\frac{1}{2}$ the first instrument is used to measure θ and with probability $\frac{1}{2}$ the second instrument is used; let z designate the resulting variable obtained by this probability mixture of x and y. Suppose also that the outcome includes an indication as to which measuring instrument is used.

The outcomes $z=x=10$ and $z=y=10$ have the same likelihood function, and in fact are the only two points with common likelihood functions. The theory of sufficient statistics then shows that these two points correspond to a single value for the minimal sufficient statistic. The sufficiency principle, which can be viewed as a derivative of the conditionality principle, then prescribes that inferences from $x=10$ should be the same as inferences from $y=10$. For present interests this is the key step in the argument from the conditionality principle to the likelihood principle.

Suppose now that we assume a meaningful position relationship between the variables and the parameters of the kind described in Section 2. The sufficiency principle would prescribe that inferences be made from $z=10$ without reference to the originating variable. This reduction clearly ignores and thereby suppresses any meaning attached to the position relationship of the originating variable to the parameter. The analysis of the discrete example thus suggests that the sufficiency principle should not be used if it suppresses or ignores any relationships beyond that embodied in the probability function.

The conditionality principle, however, if used in a natural way, seems quite appropriate here. It would lead to inferences made conditionally for the particular measuring instrument: on the basis of the statistical model for x if $z=x$ and on the basis of the statistical model for y if $z=y$. This is the kind of example used as a paradigm for the conditionality principle and it is interesting that it contains in effect a denial of the sufficiency principle.

These considerations then suggest that *when there is more than the probability relationship between variable and parameter the sufficiency and likelihood principles should be used only where they maintain or support the additional relationships.*

Most of the theory of statistical inference is concerned with making inferences about a parameter on the basis of observations on an observable variable. The thing however that is directly accessible from an observable variable is its *distribution*. The discrete variable y in Section 2 had the same distribution for $\theta=9$, 10, 11. The parameter ϕ was then used to index the distributions and it did this by recording the location of the distribution. In unbiased estimation problems arise from the fact that an unbiased estimate of a parameter using one mode of expression for the parameter will in general not be an unbiased estimate for the parameter using another mode of expression. There seem then to be good grounds, related to these and other anomolies, for claiming that *statistical inferences should be from observation to possible distribution for the observable variable.* This would leave to processes of deduction the interpretation of such inferences for a particular parametrization of the problem or mode of expression for a parameter. *A position relationship and any other relationships would then need to be expressed in terms of an observable variable and the probability distribution for that variable.*

REFERENCES

[1] Birnbaum, Allan, "On the foundations of statistical inference," *Journal of the American Statistical Association,* 57 (1962), 269–306.

[2] Cox, D. R., "Some problems connected with statistical inference," *Annals of Mathematical Statistics*, 29 (1958), 357–72.
[3] Fisher, R. A., *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd, 1956.
[4] Fraser, D. A. S., "The fiducial method and invariance," *Biometrika*, 48 (1961), 261–80.
[5] Fraser, D. A. S., "On fiducial inference," *Annals of Mathematical Statistics*, 32 (1961), 661–76.